

SYSTEM AND METHOD FOR CLASSIFYING ELECTRONICALLY POSTED DOCUMENTS

5

ABSTRACT OF THE DISCLOSURE

A method for classifying electronically posted documents includes receiving two posted documents and generating corresponding metadata summaries for each, wherein each of the metadata summaries includes at least one sub-tree structure. The structures of the two summary sub-trees within the respective metadata summaries are subsequently compared. If the two summary sub-trees are different, the two documents are deemed distinct. If the two summary sub-trees are the same, attribute values and text content of the metadata summaries are compared over a portion of the metadata summaries. If the compared attribute values and text content are determined to be the same, the documents are deemed duplicative.

201949 v01.PA (4BTP01!.DOC)
1/26/00 2:23 PM (11886.7008)